Consistency and Trust

CR MH CH

February 2, 2022

1 Progress

1.1 Model

We have:

- a 2-rounds signalling model
- a 3-rounds signalling model
- 2 equilibrium of interest
- licensing equilibrium

.

Next:

- it's enough what we have
- nice to check other equilibria
- nice to try to generalise (how though?)
- nice to check other equilibria
- possible add-ons

1.2 Experiment

We have:

- A one-player DG (Behtany's trick)
- 2 rounds
- 2 treatments

s

Next:

- review instructions
- test it, test it, test it
- pilot with 20ppl

2 Predictions

2.1 Actor

1. The more consistent in his cooperative actions the actor is, the more consistent he will be in future action.

   (a) Change the cost on the first round.

   (b) Have a same definite cost on an early round vs a later round.

2. After a betrayal/cheating, the actor has no incentives to be consistent in his cooperative behaviour again and will not be.

2.1.1 Design

- at least three rounds to show that Actor is less likely to cooperate ever after a defection irrespective of how much they cooperated before that defection.

- show consistency by varying cost to cooperate in each round

- framed as stealing/cheating to make it an obviously morally undesirable action.

2.2 Observer

1. As the actor keeps behaving consistently, the observer's trust increases.

2. The first defection reduces how moral/trustworthy they find the person much more so than the second defection.

3. In the rare case where the actor tries to rebuild trust/consistency after a betrayal/cheating, the observer will never regain full trust again.

4. Conditional trust: how does one cooperative action increase trust vs how does one betrayal/cheating action decrease trust (difference of a difference)

2.2.1 Design

- The observer may or may not know the level of temptation of the actor (= cost on each round)

- The observer can be a third party who learns the history of play or the interaction partner who would have to know previous actions.

- We need to measure the observer's level of trust

  - Partner choice (choose who to they want to be paired with to play another DG)

2

- Add a second stage and make it a trust game

- Character judgment (asking directly to rate trustworthiness or morality)

- Ask observer what how they expect the Actor to act

• We need to manipulate the history of play for the actor.

- Strategy method (ask what they expect the actor will do for every possible scenario)

- Hope enough subjects behave in all possible ways

- Use deception

3   Design

Dictator Game   We use a multi round DG reframed as a take rather give scenario to highlight the non desirability of the action. We can vary the costs or use an asymmetric token value.
Currently we have:

• dichotomous choice

• 3 rounds

• framed as taking/stealing

• value of the tokens are stochastic each round

• Perfect stranger random matching on each round

• Reciever/Observer knows doesn't know what happened in the past

Cheating task   See Mazar and Gneezy papers

• either a task in pairs where one has the opportunity to directly lie to the other

• or one does a task where he can cheat, and the other player may or may not be able to observe that

Advantage: we add principle (Lying/cheating is universally considered bad)

Trust Game   We manipulate costs and measure trust directly from money sent back
how low cost in 1st round affects cooperation in 2nd round, mediated by whether cooperate in 1st round or
how a high temptation sooner differs from a high temptation later

4 Model

4.1 Description of the signaling game

In the following we describe a simple signaling game. The game involves two players, a sender and a receiver. It consists of several rounds in which senders can decide whether or not to send a costly signal. The exact cost in a given round is determined randomly and may depend on the type of the sender. We use this model to explore under which conditions we observe equilibria that show the characteristics of principled behavior, discontinuity and complementarity. In addition, the same model will also allow us to characterize other equilibria of interest (such as equilibria with moral licensing). The game proceeds as follows.

Stage 0: Move of nature. We assume that senders can be of two different types, they are either of the superior (S) or of the inferior (I) type. The type of the sender is randomly determined before the game starts. Let p_S be the probability that a given sender is superior. Senders learn their own type, but receivers do not know the type of sender they interact with.

Stage 1–n: Signaling rounds. After the sender's type is determined, the sender faces n opportunities in which she can signal her type to the receiver. Each of these n rounds proceeds as follows:
First nature randomly determines the cost of the signal; this cost can be either low or high, $c \in \{c_l, c_h\}$ with $c_h > c_l > 0$. The probability of having a high cost is independent of the round number and of previous realized costs. However, it depends on the type of the sender; superior senders experience a high cost with probability $q_{S,h}$, whereas this probability is $q_{I,h}$ for inferior types. We use $q_{S,l} = 1 - q_{S,h}$ and $q_{I,l} = 1 - q_{I,h}$ for the probability that the respective cost is low. Moreover, we use $\bar{c}_\theta := q_{\theta,l} c_l + q_{\theta,h} c_h$ to denote the expected cost that a sender of type θ experiences in any given round. In the following we assume $q_{S,h} \leq q_{I,h}$. That is, superior types are less likely to experience a high cost. The sender (but not the receiver) learns the exact cost. Then the sender decides whether or not to pay the cost in order to send the signal.

Stage $n+1$: Receiver's decision. After the sender has made her decisions, the receiver learns in how many of the n rounds the sender paid the signaling cost. The receiver then decides whether or not to accept the sender. If the receiver rejects the sender, the payoffs in that final stage are zero for both players. If he accepts the sender, the payoffs depend on the sender's type; the sender's payoff is a_θ and the receiver's payoff is b_θ with $\theta \in \{S, I\}$. Receivers get a positive payoff from accepting superior senders, but a negative payoff from accepting inferior senders, $b_S > 0 > b_I$.

In the following, we assume the senders' strategies take the form $\sigma_\theta = \sigma_\theta(i, k, c)$. Here, the input $i \in \{1, \ldots, n\}$ is the given round for which a decision is to be made; $k \in \{0, \ldots, i-1\}$ is the number of times a signal has been sent in the past, and $c \in \{c_l, c_h\}$ is the current cost. The output takes the form $\sigma_\theta(i, k, c) \in \{0, 1\}$ where 0 refers to not sending the signal for the given input. The receiver's strategy

takes the form $\rho = \rho(k)$, where $k \in \{0, \ldots, n\}$ refers to the number of times the sender has sent the signal. Here, the output again takes values $\rho(k) \in \{0, 1\}$, where 0 means to reject the sender.

4.2 The case of two rounds

In the following, we first study the case in which there are two rounds only, $n=2$. We are interested in the following strategy profile

$$\sigma_S(i, k, c) = \begin{cases} 1 & \text{if } k = i - 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$\sigma_I(i, k, c) = \begin{cases} 1 & \text{if } (i=1 \text{ and } c=c_l) \text{ or } (i=2 \text{ and } k=1) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$\rho(k) = \begin{cases} 1 & \text{if } k = n \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

We refer to the above strategy profile as the 'zero tolerance profile'.

Claim 1. *The zero tolerance profile is a perfect Bayesian Nash equilibrium if and only if the following conditions hold:*

$$\frac{b_S}{-b_I} \geq \frac{q_{I,l}(1-p_S)}{p_S}$$

$$a_S \geq c_h + \bar{c}_S \tag{4}$$

$$c_h + \bar{c}_I \geq a_I \geq c_l + \bar{c}_I.$$

All proof are provided in the **Appendix.** Here, we give the interpretation of these conditions. The first condition states that the receiver finds it worthwhile to accept if they observe their co-player sent the signal in both rounds. The second condition states that superior senders are willing to pay the high cost even if it occurs in the first round. The last condition states that the inferior sender would be willing to pay the low cost but not the high cost.

We are also interested in the following profile:

$$\sigma_S(i, k, c) = \begin{cases} 1 & \text{if } (i=1 \text{ and } c=c_l) \text{ or } (i=2 \text{ and } k=0) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$\sigma_I(i, k, c) = \begin{cases} 1 & \text{otherwise} \\ 0 & \text{if } c=c_h \text{ or } (i=2 \text{ and } k=1) \end{cases} \tag{6}$$

$$\rho(k) = \begin{cases} 1 & \text{if } k \geq n - 1 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

We refer to the above strategy profile as the 'licensing profile'.

Claim 2. *The licensing profile is a perfect Bayesian Nash equilibrium if and only if the following conditions hold:*

$$\frac{b_S}{-b_I} \geq \frac{q_{I,l}(1-p_S)}{p_S}$$

$$a_I \leq q_{I,l}(a_I - c_l) + c_h \tag{8}$$

$$c_h \geq \bar{c}_S \geq c_l.$$

All proof are provided in the **Appendix.** Here, we give the interpretation of these conditions.

Further notes on how to continue:

1. Write the proof of Claim 1 (CR)

2. Talk about complementarity and discontinuity in the context of the model.

3. Create analogous claim for licensing equilibrium (CR)

4. Ideally, characterize all equilibria, given natural assumptions.

5. Describe some characteristics of equilibria (e.g., when does being an equilibrium plus Eq. [3] imply Eqs. [1]-[2])

4.2.1   The game

(a) Model

(1) Players: Two players: Sender & Receiver
Sender can be of two types: Superior S & Inferior I and $p^S + p^I = 1$

(2) Actions: 2 rounds in which sender can cooperate (c) or defect (d)
Afterwards, Receiver accepts or rejects.

(3) Order of moves:

	(0)	Nature determines types of Senders
Round 1	(1a)	Nature determines costs
	(1b)	Sender decides whether to cooperate
Round 2	(2a)	Nature determines costs
	(2b)	Sender decides whether to cooperate
Final stage	(3)	Receiver decides whether to accept

(4) Information that the players have:

Sender	Receiver
• Own type	• Actions of the Sender
• Current cost	
• Previous action	

(5) Payoffs:

- Costs
- Payoffs if rejected $\quad (0,0)$
- Payoff if accepted $\quad (a^S, b^S) \quad b^S > 0$
  $\qquad\qquad\qquad\quad (a^I, -b^I) \quad \;\; {}_{b^I > 0}$

(b) Assumptions

(1) Receivers only want to pair with superior Senders $\quad b^S > 0 > -b^I$

(2) Senders are generally willing to signal $\quad 2c_l > a^t$

(3) Some signalling is required $\quad p^S b^S - p^I b^I < 0$

(4) Inferior senders do not signal at all costs $\quad 2c_h < a^I$

(c) Specify players' strategies

$$
\begin{aligned}
\sigma^S &= (\sigma_h^{S1}, \sigma_l^{S1}, \sigma_{hc}^{S2}, \sigma_{hd}^{S2}, \sigma_{lc}^{S2}, \sigma_{ld}^{S2}) \\
\sigma^I &= (\sigma_h^{I1}, \sigma_l^{I1}, \sigma_{hc}^{I2}, \sigma_{hd}^{I2}, \sigma_{lc}^{I2}, \sigma_{ld}^{I2}) \\
\rho &= (\rho_{cc}, \rho_{cd}, \rho_{dc}, \rho_{dd})
\end{aligned}
\tag{9}
$$

where S, I is the Sender's type, $1, 2$ the round number, h, l the cost on that round, and c, d the action on the previous round.

(d) Properties:

(1) Threshold

(2) Betweenness

(3) Complementarity $\quad \sigma_{(c,..,c)}^{\theta,n+1} > \sigma_{(c,..,c)}^{\theta,n} \, \forall n$

(4) Discontinuity $\quad \sigma_h^{\theta,n} = \underline{c} \quad \forall h : D \in h$

(5) Existence of committed type $\quad \sigma^{S,n} = \bar{c}$

(6) licensing (inverse of complementarity) $\quad \sigma_h^{\theta,n} < \sigma_{(d,...,d)}^{\theta,n} \quad h$ such that $c \in h$

(7) Anti-complementarity $\quad \exists n \quad \sigma_{(c,..,c)}^{\theta,n+1} < \sigma_{(c,..,c)}^{\theta,n}$

(8) Monotonicity in types $\quad \sigma_h^{S,n} > \sigma_h^{I,n}$

4.3 Signalling Game with 3 rounds

4.3.1 The game

Equivalent to the 2 rounds version but the Sender gets an additional thrid round to send the signal or not. Players, Actions, Informations and Payoffs stay the same.

(a) New strategies

$$\sigma^S = (\sigma_h^{S1}, \sigma_l^{S1}, \sigma_{hc}^{S2}, \sigma_{hd}^{S2}, \sigma_{lc}^{S2}, \sigma_{ld}^{S2}, \sigma_{hcc}^{S3}, \sigma_{hcd}^{S3}, \sigma_{hdc}^{S3}, \sigma_{hdd}^{S3}, \sigma_{lcc}^{S3}, \sigma_{lcd}^{S3}, , \sigma_{ldc}^{S3}, \sigma_{ldd}^{S3})$$
$$\sigma^I = (\sigma_h^{I1}, \sigma_l^{I1}, \sigma_{hc}^{I2}, \sigma_{hd}^{I2}, \sigma_{lc}^{I2}, \sigma_{ld}^{I2}, \sigma_{hcc}^{I3}, \sigma_{hcd}^{I3}, \sigma_{hdc}^{I3}, \sigma_{hdd}^{I3}, \sigma_{lcc}^{I3}, \sigma_{lcd}^{I3}, , \sigma_{ldc}^{I3}, \sigma_{ldd}^{I3}) \tag{10}$$
$$\rho = (\rho_{ccc}, \rho_{ccd}, \rho_{cdc}, \rho_{cdd}, \rho_{dcc}, \rho_{dcd}, \rho_{ddc}, \rho_{ddd})$$

where S, I is the Sender's type, $1, 2, 3$ the round number, h, l the cost on that round, and c, d the action on the two previous rounds.

4.3.2 Equilibrium analysis

1st equilibrium of interest The superior type cooperates on the first round regardless of costs, and then cooperates in subsequent rounds as long as he cooperated in all previous rounds. If he defected once, he does not cooperate. The Inferior type cooperates on the first round only when he sees a low cost. He cooperates on subsequent round in the same fashion as the superior type. The Receiver accept only and only if there are no defections at all.

$$\sigma^S = (1, 1, 1, 0, 1, 0, 1, 0, 0, 0)$$
$$\sigma^I = (0, 1, 1, 0, 1, 0, 1, 0, 0, 0) \tag{11}$$
$$\rho = (1, 0, 0, 0, 0, 0, 0, 0)$$

We get:
Receiver:

- Receiver after ccc $\quad \rightarrow \quad accept \quad \rightarrow \quad \frac{b^S}{b^I} \geq \frac{q_l^I(1-p_S)}{P_S}$

- Receiver after $ccd, cdc, dcc, cdd, dcd, ddc, ddd \quad \rightarrow \quad reject \quad \rightarrow \quad -b^I \geq 0$

Superior & Inferior Sender on Round 3:

- Sender after cc in $r_{1,2}$ and C_h now $\quad \rightarrow \quad cooperate \quad \rightarrow \quad C_h \leq a^S$

- Sender after cc in $r_{1,2}$ and C_l now $\quad \rightarrow \quad cooperate \quad \rightarrow \quad C_l \leq a^S$

- Sender after cd, dc, dd in $r_{1,2}$ and C_h now $\quad \rightarrow \quad defect \quad \rightarrow \quad 0 \geq C_h$

- Sender after cd, dc, dd in $r_{1,2}$ and C_l now $\quad \rightarrow \quad defect \quad \rightarrow \quad 0 \geq C_l$

Superior & Inferior Sender on Round 2:

- Sender after d in r_1 and C_h now $\quad \rightarrow \quad defect \quad \rightarrow \quad C_h \leq 0$

- Sender after d in r_1 and C_l now $\quad \rightarrow \quad defect \quad \rightarrow \quad C_l \leq 0$

- Sender after c in r_1 and C_h now $\quad \rightarrow \quad cooperate \quad \rightarrow \quad a^S \geq C_h q_h^S + C_l q_l^S + C_h$

- Sender after c in r_1 and C_l now $\quad \rightarrow \quad cooperate \quad \rightarrow \quad a^S \geq C_h q_h^S + C_l q_l^S + C_l$

Superior Sender on Round 1:

- Sender after C_h now $\quad \rightarrow \quad cooperate \quad \rightarrow \quad a^S \geq C_h q_h^S + C_l q_l^S + C_h q_h^S + C_l q_l^S + C_h$

- Sender after C_l now $\quad \rightarrow \quad cooperate \quad \rightarrow \quad a^S \geq C_h q_h^S + C_l q_l^S + C_h q_h^S + C_l q_l^S + C_l$

Inferior Sender on Round 1:

- Sender after C_h now $\quad \rightarrow \quad defect \quad \rightarrow \quad a^I \leq C_h q_h^I + C_l q_l^I + C_h q_h^I + C_l q_l^I + C_h$

- Sender after C_l now $\quad \rightarrow \quad cooperate \quad \rightarrow \quad a^I \geq C_h q_h^I + C_l q_l^I + C_h q_h^I + C_l q_l^I + C_l$

2nd equilibrium of interest   Same as above for the Superior type and the Receiver. But now the inferior type may cooperate in the second round even after a defection in the first round.

$$
\begin{aligned}
\sigma^S &= (1, 1, 1, 0, 1, 0, 1, 0, 0, 0) \\
\sigma^I &= (0, 1, 1, 0, 1, 0, 1, 0, 0, 0) \\
\rho &= (1, 0, 0, 0, 0, 0, 0, 0)
\end{aligned}
\tag{12}
$$

5 Appendices

5.1 Proofs

Claim 1   zero-tolerance equilibrium:
Receiver:

- Receiver after cc $\rightarrow$ $accept$ $\rightarrow$ $\frac{b^S}{b^I} \geq \frac{q_l^I(1-p_S)}{p_S}$

- Receiver after cd, dc, dd $\rightarrow$ $reject$ $\rightarrow$ $-b^I \geq 0$

Superior & Inferior Sender on Round 2:

- Sender after d in r_1 and C_h now $\rightarrow$ $defect$ $\rightarrow$ $C_h \leq 0$

- Sender after d in r_1 and C_l now $\rightarrow$ $defect$ $\rightarrow$ $C_l \leq 0$

- Sender after c in r_1 and C_h now $\rightarrow$ $cooperate$ $\rightarrow$ $a^S \geq C_h$

- Sender after c in r_1 and C_l now $\rightarrow$ $cooperate$ $\rightarrow$ $a^S \geq C_l$

Superior Sender on Round 1:

- Sender after C_h now $\rightarrow$ $cooperate$ $\rightarrow$ $a^S \geq C_h q_h^S + C_l q_l^S + C_h$

- Sender after C_l now $\rightarrow$ $cooperate$ $\rightarrow$ $a^S \geq C_h q_h^S + C_l q_l^S + C_l$

Inferior Sender on Round 1:

- Sender after C_h now $\rightarrow$ $defect$ $\rightarrow$ $a^I \leq C_h q_h^I + C_l q_l^I + C_h$

- Sender after C_l now $\rightarrow$ $cooperate$ $\rightarrow$ $a^I \geq C_h q_h^I + C_l q_l^I + C_l$

Claim 2 licensing equilibrium:
Receiver:

- Receiver after cc $\rightarrow$ Out of equilibrium!

- Receiver after cd, dc $\rightarrow$ $accept$ $\rightarrow$???

- Receiver after dd $\rightarrow$ $reject$ $\rightarrow$ $-b^I \geq 0$

Superior Sender on Round 2:

- Sender after c in r_1 and C_h now $\rightarrow$ $defect$ $\rightarrow$ $C_h \leq 0$

- Sender after c in r_1 and C_l now $\rightarrow$ $defect$ $\rightarrow$ $C_l \leq 0$

- Sender after d in r_1 and C_h now $\rightarrow$ $cooperate$ $\rightarrow$ $a^S \leq C_h$

- Sender after d in r_1 and C_l now $\rightarrow$ $cooperate$ $\rightarrow$ $a^S \leq C_l$

Inferior Sender on Round 2:

- Sender after d in r_1 and C_l now $\rightarrow$ $defect$ $\rightarrow$ $a^I \geq C_h$

Superior Sender on Round 1:

- Sender after C_h now $\rightarrow$ $defect$ $\rightarrow$ $C_h \geq -C_h q_h^S - C_l q_l^S$

- Sender after C_l now $\rightarrow$ $cooperate$ $\rightarrow$ $C_l \leq -C_h q_h^S - C_l q_l^S$

Inferior Sender on Round 1:

- Sender after C_h now $\rightarrow$ $defect$ $\rightarrow$ $a^I \geq q_l^I(-C_l + a^I) + C_h$

- Sender after C_l now $\rightarrow$ $cooperate$ $\rightarrow$ $C_l \leq -C_h q_h^I - C_l q_l^I$

6 oTree Screenshots

Introduction

Thank you for taking part in this study!

In this study, you are about to engage in 2 interactions with a **different participant** in each. Every participant taking part in this study will be given the role of either **decider** or **receiver**. This role is fixed throughout the whole study. You will keep the same in every interaction. In each interaction, one decider and one receiver are paired up. Only the decider, not the receiver, will have a decision to make.

In each interaction, there is the opportunity to earn a bonus payment. Both the bonus of the decider and the receiver depend on the decisions of the decider. At the end of the study, all the bonus you have earned from every interaction will be summed up and paid to you after completing the study.

You must stay on task at all times. Otherwise you will automatically be disconnected and lose your bonus.

Please answer the following questions to continue:

With how many participants will you be playing in this study?

○ 0 other participants
○ 1 other participant
○ Different participants

What will your bonus payment depend on?

○ There is no bonus possible in this study.
○ My bonus payment depends on luck.
○ My bonus payment depends on a decision taken by one of the participants.

Next

Instructions

You are the **Decider**. This means that in every interaction you have a decision to make. In each interaction, both you and the receiver are allocated with **£1.00** each. You have the opportunity to take the money of the receiver or leave it. If you decide to take it, you will earn **£2.00** in that interaction and the receiver gets nothing. If you decide to leave it, you keep your allocated **£1.00** and so does the receiver. The receiver will know his final amount, but not who decided.
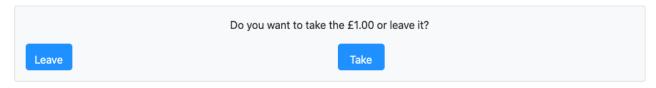
Remember, in each interaction you are paired up with someone different and make a new decision. What happened previously is in the past and you are allocated £1.00 again.

Below you can see what your choices in each interaction will look like.

Interaction 1

Both you and the receiver have been endowed with **£1.00** each. You have the opportunity to take the money from the receiver, or to leave it.

- If you take it, your total payoff for this interaction will be **£2.00**.
- If you leave it, your total payoff for this interaction will be **£1.00**.

Do you want to take the £1.00 or leave it?

Leave   Take

Please answer the following questions to continue:

Who makes the decision in each interaction?

○ You, the decider.
○ The receiver.
○ Both the decider and the receiver.

What will be your total payoff in this round if you choose to take the £1.00?

○ £0.
○ £1.00.
○ £2.00.

Next

Your Decision

Interaction 1

Both you and the receiver have been endowed with **£3.00** each. You have the opportunity to take the money from the receiver, or to leave it.

- If you take it, your total payoff for this interaction will be **£6.00**.
- If you leave it, your total payoff for this interaction will be **£3.00**.

Do you want to take the £3.00 or leave it?

Leave   Take

Your Decision

Interaction 1

Both you and the receiver have been endowed with **£1.00** each. You have the opportunity to take the money from the receiver, or to leave it.

- If you take it, your total payoff for this interaction will be **£2.00**.
- If you leave it, your total payoff for this interaction will be **£1.00**.

Do you want to take the £1.00 or leave it?

Leave   Take

Results

You decided to leave the £3.00 of the receiver. Your payoff in this round is of **£3.00**.

Next

Results

You decided to take the £3.00 of the receiver. Your payoff in this round is of **£6.00**.

Next

The End

The task is now over. Here are your final results: In total across all rounds, you got **£9.00**.

Round	Bonus	Total
1	£3.00	
2	£6.00	£9.00

Next

Quick Survey

We appreciate feedback. If you have any comments on the implementation of this study, or on the user interface, please write it down below.

Would you accept to be matched with another participant in order to be the receiver of another decider? tYou would receive the bonus for the receiver from their decisions. It would be at no cost to you, the payment would be directly made in Prolific at a later time.

○ Yes
○ No

Next

Payment

The study is now over. **Thank you for participating!** Below is your total score and the money you earned.

Total of the bonus in GBP   £9.00
Participation payment   £2.00

Final Payment to you (participation payment + bonus)	£11.00
Click on the **next button** to be redirected to Prolific and claim your payment.	
	Next